Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

1g84Mr
cop. 3

LIBRARY
CURRENT SERIAL RECORD
JUL 13 1954
U. S. DEPARTMENT OF AGRICULTURE

Sampling Routines and the Accuracy of Patrons' Butterfat Tests

Marketing Research Report No. 66

U. S. DEPARTMENT OF AGRICULTURE
AGRICULTURAL MARKETING SERVICE
MARKETING RESEARCH DIVISION

Washington, D. C.

May 1954

ACKNOWLEDGMENTS

The data on which this report is based were made available by the St. Louis Milk Market Administrator and the Departments of Dairy Husbandry and Agricultural Economics at the University of Missouri. Participating in the development of plans for gathering the data were: Fred Shipley, St. Louis Milk Market Administrator; A. C. Ragsdale and J. E. Edmonson of the Department of Dairy Husbandry; and Ben Frame and E. H. Matzen of the Department of Agricultural Economics of the University of Missouri.

The sampling and testing work was done by testers employed by the St. Louis Milk Market Administrator. Procedures were specified by Mr. Edmonson, who also made spot checks of the testing work.

The analysis of the data and the preparation of this report were done under authority of the Agricultural Marketing Act of 1946 (RMA, Title II).

CONTENTS

	Page
Summary	i
Introduction	1
Results	4
Variability of daily tests	4
Factors that affect variability of tests	7
Compromising plant tests and check tests	10
Composite samples	11
Comparability of fresh and composite samples	13
Simple vs. weighted averages of daily tests	15
Conclusions	15
Literature cited	16
Appendix	18

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price 25 cents

SUMMARY

Milk deliveries of 194 producers during one month were analyzed to ascertain how closely the tests of composite samples, or of specified numbers of random fresh samples, approached each producer's true average butterfat percentage for the month. Dairy plant operators and officials of cooperatives and governmental agencies need such information for comparing and choosing among various routines for the sampling and butterfat testing of milk delivered by patrons of dairy plants.

It was found that the butterfat percentages of about 2 daily samples in 3 fell within 0.196 (one standard deviation) of the producer's monthly average. Outside temperature on the day preceding the taking of the sample, day of the week, and week of the month were related to the butterfat percentage, but the relation was not sufficiently strong to be of practical importance. More frequent sampling of milk from some producers would increase the overall accuracy of fresh-sample testing appreciably. The additional samples could be taken from deliveries from small herds, high-testing herds, or herds having large day-to-day variations in quantities delivered.

Suggestions are made for statistical methods to determine whether a testing program is properly conducted. Such methods used as a supplement to a moderately intensive check-testing program could reduce testing costs and permit more careful supervision of testing at those plants at which the estimates of producers' butterfat percentages are found to be least accurate.

Composite samples gave lower monthly average butterfat percentages than samples of fresh milk tested daily. The monthly average percentages of butterfat from composites tested every 10th day were lower than the weighted averages of daily tests by 0.04, the standard deviation of the differences for individual producers being 0.045. Composites tested every 15th day gave an average percentage lower by 0.06, with a standard deviation of the individual differences of 0.053.

Although tests from composite samples averaged low, they were consistent. About two-thirds of the butterfat percentages from 10-day composite samples could be expected to fall within -0.084 and +0.006 of the true test, a range of 0.090. It required about 12 daily samples to give average tests of which two-thirds lay within this narrow a range. Fifteen-day composites were equivalent to 10 daily samples on this basis.

With as few as 5 daily samples a month, about 96 producers in 100 in any month would receive a test as high or higher than with 10-day composite samples. Three fresh samples would be equivalent to 15-day

composites on the basis of equal or greater advantage to about 96 producers in 100.

Simple averages of the daily butterfat percentages were nearly identical with weighted averages, differing by less than 0.005 in 129 of 194 cases.

As the data on which this study is based are only from one month and from one plant, the conclusions may differ from those found by similar studies in other seasons and at other plants.

SAMPLING ROUTINES AND THE ACCURACY OF PATRONS' BUTTERFAT TESTS

By Louis F. Herrmann, W. G. Bryan, and Elsie Anderson
Agricultural Marketing Service
Market Organization and Costs Branch

INTRODUCTION

The taking of a sample that will accurately represent the butterfat content of the milk in a single day's delivery by one producer is a problem to which much consideration has been given (10, 14, 18) 1/. Yet, even if each day's delivery is sampled satisfactorily, there remains the problem of accurately representing all the milk delivered during a month in order to establish a basis of payment which will be equally fair to plants and their patrons.

Fresh samples tested daily may be considered the ideal measure of the butterfat in a month's deliveries. The "true" average percentage for the month would be the average of daily percentages weighted by the quantity of milk delivered daily. But less costly sampling and testing routines are needed.

A common alternative is to sample each producer's delivery daily, combining daily samples into 10- to 15-day composites. This requires only 2 or 3 tests a month, but a sampler must be present each day when the milk is received.

It has been demonstrated that composite samples give lower butterfat percentages than do fresh samples. The differences are small and composite samples are considered acceptable if properly preserved. Some published results of studies of butterfat percentages found in composite samples are shown in table 1. Wilster and Robichaux (17) reported that the average difference in 60 comparisons was not statistically significant; but in view of the general agreement among the different studies there can be little doubt that in a larger number of comparisons the differences would be found to be significant.

While the average of butterfat percentages of a large number of composite samples will be lower than the average of the fresh samples, it is important to know the proportion of composite samples that will have the lower percentage. Composites were found to be lower in 78 percent of 290 comparisons by Monroe (12) and in 60 percent of 153 comparisons by Meade and Leckie (11). It would be helpful, too, to know the proportion of the butterfat percentages of composite samples that differ by less than stated amounts from the average of daily fresh samples, say, within 0.1 or 0.2, but none of the reports reviewed gave such data.

1/ Underscored numbers in parentheses indicate references to Literature Cited, p. 16.

Table 1.--Butterfat percentages determined from composite samples compared with averages of daily tests as reported by various investigators

Source	Number		Butterfat percentages		
	Comparisons	Days in composite	Average of composites	Average of daily fresh tests	Difference: Composite - daily
Campbell (1)	100	15	3.86	3.89	-0.03
	99	15	3.74	3.75	-0.01
England and D'Ambrogi (3)	-	15	1/ 4.28	4.36	-0.08
	-	15	2/ 4.19	4.36	-0.17
Farrington (5)	20	7	3/ 3.96	3.91	+0.05
	20	7	4/ 3.93	3.91	+0.02
Hunziker (6)	350	14	3/ 4.87	4.90	-0.03
	350	14	4/ 4.86	4.90	-0.04
Judkins (7)	48	15	3/ 4.06	4.00	+0.06
	48	15	4/ 4.03	4.00	+0.03
Kent (8)	52	13	3/ 3.95	4.01	-0.06
	52	13	4/ 4.00	4.01	-0.01
Lucas (9)	231	15	-	-	-0.11
Meade and Leckie (11)	153	10	4.10	5/ 4.19	-0.09
Monroe (12)	290	7	5.13	5.22	-0.09
Potts 6/	60	7	3/ 4.35	4.39	-0.04
	60	7	4/ 4.36	4.39	-0.03
	60	7	7/ 4.38	4.39	-0.01
	60	14	3/ 4.04	4.13	-0.09
	60	14	4/ 4.00	4.13	-0.13
	60	14	7/ 4.04	4.13	-0.09
Tracy and Tuckey (16)	348	7	4.64	4.70	-0.06
	343	7	4.71	4.74	-0.03
	100	7	4.13	4.15	-0.02
	50	7	4.07	4.10	-0.03
Wilster and Robichaux (17)	60	5	4.75	4.77	-0.02
	60	7	4.51	4.53	-0.02
	60	15	4.52	4.55	-0.03

1/ Samples held at 45° and 65° F.
2/ Samples held at 80° and 100° F.
3/ Composite of equal daily parts.
4/ Composite of aliquot daily parts.
5/ Average of three fresh samples during the month.
6/ Data from an unpublished M. S. thesis, Cornell University, as cited by Judkins (7).
7/ Samples taken with McKay sampler.

An alternative to composite sampling is to sample deliveries on one or more days, test the samples individually, and consider the average to be representative of the month's deliveries. This does not require as much labor for sampling and testing, but the resulting average percentages differ more or less from the true average. The differences are mainly random; they are distributed equally above and below the true average. However, there may be systematic differences or biases. For example, samples taken on the same day of the week throughout the month might give consistently low or high tests.

Several investigators have found that tests of fresh samples taken on 3 or more days throughout a month will be closer to the average of daily tests than will a composite (2, 6, 7, 9, 13). Teal (15) found that annual butterfat receipts computed from tests of 5 fresh samples monthly for each producer at milk-receiving stations, compared closely with the butterfat determined by tests of samples taken from the tankloads shipped out. Erdmann (4) concluded that butterfat tests based on random samples of fresh milk stratified as to week of the month, were valid for verifying the milk tests of handlers. Despite these results, there has been much distrust of the use of fresh-milk samples as a basis for payment. In some States only tests of composite samples are a legal basis for paying producers. This distrust is due partly to the risk of manipulation by producers, but it is also due partly to the fact that experimental results have not been presented in forms most appropriate for judging their usefulness.

The object of the study reported here was to estimate how closely an individual producer's monthly weighted average butterfat percentage is represented by tests of 10- and 15-day composite samples or by samples taken and tested on one or more days during the month. This information would help to reconcile the conflict between cost of sampling and testing and the desire for less variable results.

Samples were taken daily from the milk of 194 producers delivering to one plant in St. Louis, Mo., during April 1948. The data were analyzed statistically, both by simple tabulation and by methods which made it possible to measure more precisely the effect of the different factors that caused variations in the daily tests. Most of the analyses were carried out with data for 192 of the producers, and for 28 days of the month, in order to permit dividing the producers into 6 and 8 groups of equal size, and the days into 4 weeks. 2/

2/ For details of the sampling and testing procedure and the methods of statistical analysis, see appendix, pp. 18-23.

RESULTS

Variability of Daily Tests

The daily tests of individual producers differed from the weighted averages of 28 days' tests according to the pattern shown in figure 1. About three-fourths of the differences were between -0.2 and +0.2. An analysis of variance of the daily tests revealed that for the average producer 2 in 3 daily tests fell within -0.196 and +0.196 of the simple average of 28 tests. This leads to the expectation that when a producer's milk is tested on one day during a month the chances are 2 in 3 that the result will be within 0.192 of the simple average fat percentage of tests made on each day of the month.

This result contains enough information to permit determination of the precision with which a producer's monthly average test can be estimated from the average of tests on 2, 3, 4, or more days selected at random during the month. This makes it possible to ascertain how many daily tests are necessary in order to obtain average butterfat percentages, a specified proportion of which will fall within given limits of the true averages. For example, 4 daily tests a month can be expected to yield average butterfat percentages, 19 in 20 of which will fall within 0.2 of the true fat percentages. The limits for agreement on 2 in 3, 19 in 20, and 99 in 100 averages of various numbers of daily tests are given in table 2. 3/

In table 2, the limits within which 2 in 3 averages may be expected to vary from the true monthly average are shown as "found" and as "calculated." The found limits were based on the tests of selected days--those of the 15th and 30th for the average of 2 tests, the 10th, 20th, and 30th for 3, and so on. The method used for measuring the found variability is simple, but it uses only part of the information gathered in the study. The calculated limits, which are based on all the data, indicate more reliably the limits that would be found in repeated studies.

For specified numbers of fresh samples the variations of the average tests about each producer's true average test are indicated in table 2. The differences found fluctuated around zero. The average differences in butterfat percentages were as follows:

3/ The equation for the standard error of the average (mean) of two or more tests is $\sigma_M = \sigma \sqrt{(N-n)/(nN)}$, where σ_M is the standard error of the mean, σ the standard deviation of individual tests, N the number of days in the month, and n the number of daily tests averaged.

301087 O - 54 - 2

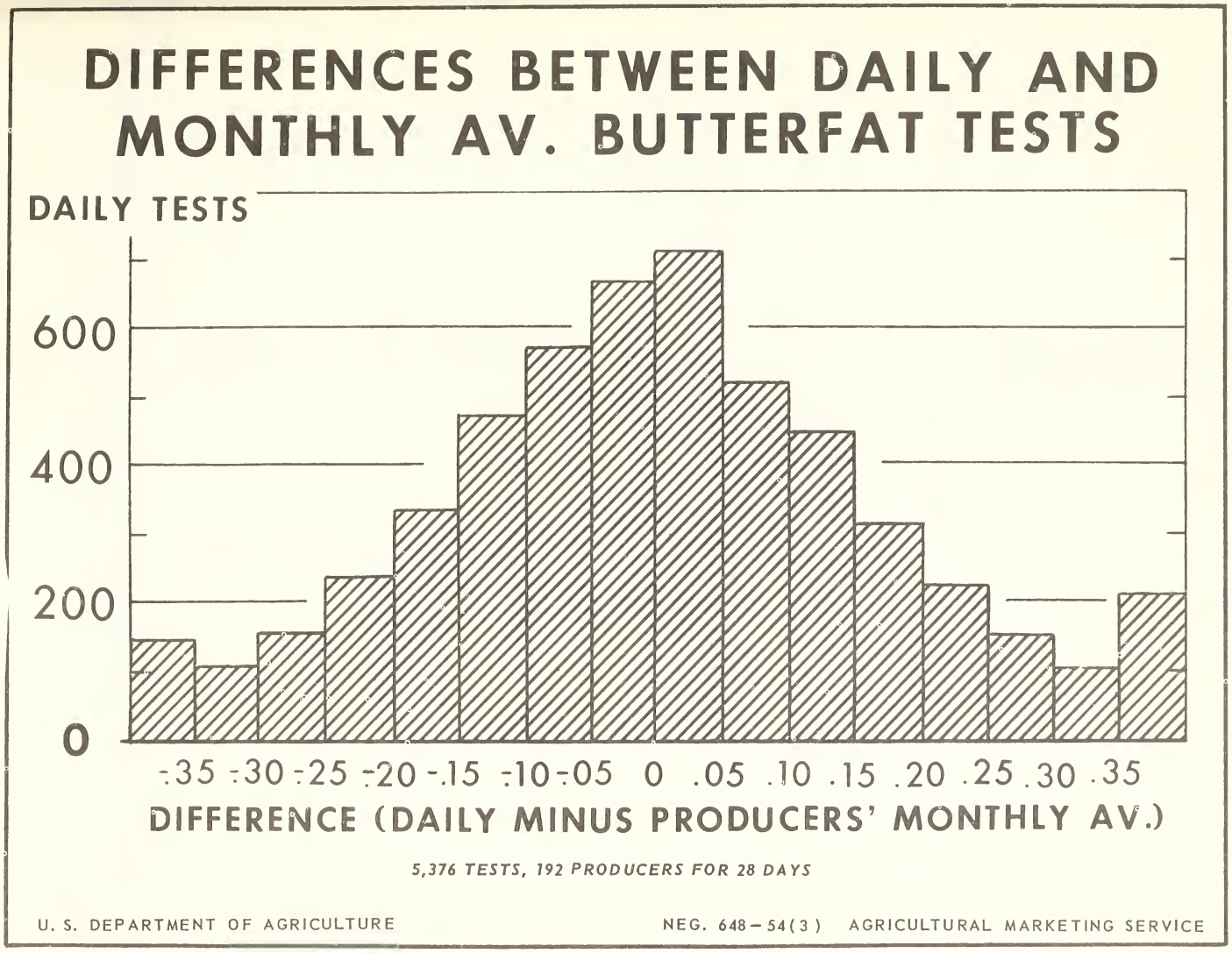
DIFFERENCES BETWEEN DAILY AND MONTHLY AV. BUTTERFAT TESTS
DAILY TESTS
600
400
200
0
-.35 -.30 -.25 -.20 -.15 -.10 -.05 0 .05 .10 .15 .20 .25 .30 .35
DIFFERENCE (DAILY MINUS PRODUCERS' MONTHLY AV.)
5,376 TESTS, 192 PRODUCERS FOR 28 DAYS
U. S. DEPARTMENT OF AGRICULTURE
NEG. 648-54(3) AGRICULTURAL MARKETING SERVICE

Figure 1.--Frequency distribution of 5,376 differences between daily tests and producers monthly weighted average tests, 192 producers for 28 days.

Number of daily tests	Average difference (found minus true)
	Percent
2	0.012
3	0.008
4	0.005
5	-0.020
6	-0.006
7	-0.020
10	0.000
15	-0.002

The differences were statistically significantly different from zero only for 5 and 7 daily tests.

Table 2.--Expected variation of the average butterfat percentages of specified numbers of random daily tests from the true monthly average

Number of daily tests	Limits of expected variation from true average			
	2 in 3 averages		19 in 20 averages	99 in 100 averages
	Found 1/	Calculated		
	Percent	Percent	Percent	Percent
1	-	0.192	0.376	0.495
2	0.119	0.134	0.263	0.346
3	0.095	0.107	0.210	0.276
4	0.087	0.091	0.178	0.235
5	0.077	0.080	0.157	0.206
6	0.058	0.071	0.139	0.183
7	0.055	0.065	0.127	0.168
8	-	0.059	0.116	0.152
9	-	0.055	0.108	0.142
10	0.046	0.051	0.100	0.132
11	-	0.047	0.092	0.121
12	-	0.044	0.086	0.114
13	-	0.041	0.080	0.106
14	-	0.038	0.074	0.098
15	0.032	0.036	0.071	0.093

1/ The "found" variation was based on the average of the tests on the 15th and 30th for 2 tests a month, the 10th, 20th, and 30th for 3 tests, etc.

Factors That Affect Variability of Tests

The variations shown in table 2 are what may be expected if samples are taken at random during the month, and if each producer's deliveries show the same pattern of variation in percentages of fat from day to day. But it is possible that the day-to-day variations in butterfat tests differ (1) between large and small herds, (2) between herds with a high average butterfat percentage and herds with a low butterfat average, or (3) between herds with large day-to-day variations in the quantity of milk delivered and herds with small daily variations in milk deliveries. Or, it may happen that (1) tests on a given day of the week tend to be high or low, (2) tests in one week average higher than tests in another, or (3) the temperature may influence the test.

If such factors are important causes of variation in fat tests, and if their effects could be allowed for in the kind of sampling routine to be followed, or in an adjustment to the test, test averages would be more accurate.

In the study reported here it was found that three of the factors listed above ranked in the following order, according to the degree of their relationship to the variance of butterfat tests:

Temperature
Day of week
Week of the month

For each of 4 temperature groups, producers' average butterfat percentages were as follows:

Temperature group Degrees F.	Average butterfat test Percent
69-76	3.43
61-68	3.46
51-60	3.49
44-50	3.49
General average	3.468

The differences among the average tests for the different temperature groups were statistically significant at the 1-percent level or higher.

Producers' butterfat tests averaged for each day of the week were as follows:

Day of week	Average butterfat test Percent
Sunday	3.45
Monday	3.44
Tuesday	3.47
Wednesday	3.47
Thursday	3.49
Friday	3.49
Saturday	3.47
General average	3.468

The percentages for Sunday and Monday were significantly lower than those for the other days of the week.

As related to week of the month producers' butterfat percentages were as follows:

Week of month	Average butterfat test Percent
1	3.46
2	3.49
3	3.47
4	3.45
General average	3.468

The differences among the 4 weekly averages were statistically significant at the 1-percent level, but not when allowance was made for the variation traceable to temperature on the day preceding that of the test.

Making allowance for the variation in fat percentages traceable to temperature groups would reduce the within-producer standard deviation by 0.0018. That is, with adjustment for temperature, two-thirds of all daily tests would fall within 0.1939 of the producers' monthly average fat percentages, as compared with a limit of 0.1957 for unadjusted data (appendix table 6). Adjustments for day of the week on which the sample was taken, and for week of the month would have less than one-third the effect of temperature, reducing the within-producer standard deviation by 0.0005. Furthermore, it was found that when differences in temperature were accounted for, the remaining week-to-week changes in butterfat percentages were not statistically significant.

An allowance for temperature would reduce the variability of butterfat tests by 0.0018. But by increasing the number of samples per month to 5 instead of 4, the variability is reduced by 0.011, or about 6 times as much.

There are some categories of producers whose butterfat tests are a great deal more variable than others. The accuracy of tests in these

categories would be increased appreciably by taking 1 or 2 more samples per month. When producers were grouped on the basis of total pounds of milk delivered during the month, it was found that 3 samples per month for producers who delivered 11,700 pounds or more gave results with as little variation as 4 tests for the average producer (table 3). But producers who delivered less than 11,700 pounds for the month required 5 samples to come within the limits of variability of 4 tests for the average producer.

When producers were classified according to average butterfat test for the month, 4 tests gave average variability for producers averaging up to 3.50 percent butterfat, and 5 tests were needed for producers averaging more than 3.50 percent.

Quantities of milk delivered by some producers fluctuated very little from day to day, while deliveries of others were more variable. Three samples gave average variability of butterfat tests for producers whose coefficient of variability of milk deliveries was less than 7.70 percent; from 7.70 and up it required 5 samples.

Table 3.--Expected variation of the mean of specified numbers of daily tests from the true monthly average; producers grouped by size, average butterfat test, and variability of milk deliveries

Number of tests	Limits of expected differences from true average in 2 of 3 means: producers grouped according to -					
	Size 1/		Average butterfat test 2/		Variability of milk deliveries 3/	
	Small	Large	Low	High	Even	Uneven
2	0.140	0.112	0.123	0.149	0.115	0.150
3	0.112	0.090	0.099	0.120	0.092	0.120
4	0.095	0.077	0.084	0.102	0.078	0.102
5	0.084	0.067	0.074	0.089	0.069	0.090
6	0.075	0.060	0.066	0.080	0.061	0.080
7	0.068	0.054	0.060	0.072	0.056	0.073

1/ Size measured in pounds of milk delivered during the month: Small producers delivering less than 11,700 pounds; large producers delivering more than 11,700 pounds.

2/ Low butterfat tests were monthly average tests of 3.5 percent or less; high tests were more than 3.5 percent.

3/ Even producers had a coefficient of variation of daily quantities of milk delivered of 7.70 percent and under; uneven producers, more than 7.70 percent.

With smaller groupings of producers, differences in the number of samples required for average accuracy were even greater than shown in table 3. Six samples would be needed for producers testing more than

3.76 percent butterfat. Only 3 would be needed for producers whose butterfat tests averaged less than 3.34 percent, or whose daily deliveries had a coefficient of variation of less than 6.12.

Variability of deliveries tended to be greater in small than in large herds. It was not possible with the data available to learn the relative advantage of more frequent sampling of milk delivered by small producers over that of more frequent sampling of milk from producers whose deliveries fluctuated considerably from day to day.

Compromising Plant Tests and Check Tests

The results of this study may also be considered from the standpoint of rules for compromising differences between tests made by the plant operator and check tests. A common rule is to compare the check test and the plant test for each producer, and to use the check test rather than the plant test when the two average fat percentages differ by 0.2 or more. Suppose that the check test is based on 4 fresh tests, and the plant test is based, on a fresh test each day. In this case, it is to be expected that 1 in 36 check tests will be less accurate than the plant test. Thus, there is no statistical basis for concluding that there has been any misrepresentation in the plant tests if no more than 1 in 36 pairs of average tests differs by more than 0.2.

If the plant test is also based on 4 daily tests, differences between the two average fat percentages of 0.2 or more would occur with greater frequency--about 1 in 8 tests. But the average of 2 such 4-day averages would be highly accurate, and random differences of 0.2 or more from the true monthly average would occur on the average no more than once in 1,700 tests (provided each of the 8 tests had been made on a different day). Thus, it would be valid to follow the practice of averaging plant and check tests to get the basis of payment in such a case. But averaging is recommended only when the frequency of differences of a given size is in reasonable agreement with the frequency to be expected by chance.

A predictable proportion of differences exceeding any given limit--say, differences greater than 0.2--could arise by chance between any two series of tests conducted with scrupulous regard for accuracy. This affords some basis for determining whether systematic errors are being made in a series of tests. If such errors are being made, they would show up as an excessive proportion of averages exceeding specified differences. For example, differences of more than 0.2 occurring in more than 1 in 8 tests would indicate a bias in the plant tests, the check tests, or both, if both are based on 4 fresh samples.

Such a measure of bias is more precise if producers are grouped by size, the frequency of "outside" differences being tabulated separately for small and large producers. As shown above, when small and large producers are sampled the same number of times, the results of two sets of samples from large producers will agree more closely than will results from small producers.

Expected percentage distributions of differences between two series of averages are shown in table 4. It is unlikely that any series of tests would correspond exactly to the distributions shown in this table, but the agreement can be tested by the X^2 (chi-square) test.

It might be desirable to survey the testing program at each milk plant periodically by the method described above to determine whether any bias is present. If a bias is shown to be present, steps can be taken to improve the sampling and testing methods; the number of fresh samples tested for each producer can be increased, or the rules for compromising differences can be made more rigorous.

Table 4.--Proportion of differences of specified sizes to be expected when comparing two series of butterfat averages, each based on four fresh tests, for various groups of producers

Size of difference	All producers	Producers grouped according to -					
		Size 1/		Average butterfat test 2/		Variability of milk deliveries 3/	
		Small	Large	Low	High	Even	Uneven
Percent of butterfat	Percent	Percent	Percent	Percent	Percent	Percent	Percent
0.00 - 0.05	30	29	35	33	28	35	27
0.06 - 0.10	26	25	29	27	24	28	24
0.11 - 0.15	19	19	19	19	19	19	19
0.16 - 0.20	13	13	10	12	13	10	13
0.21 - 0.25	7	7	4	6	8	5	8
0.26 - 0.30	3	4	2	2	4	2	5
0.31 and over	2	3	1	1	4	1	4
Total	100	100	100	100	100	100	100

1/ Size measured in pounds of milk delivered during the month: Small producers delivering less than 11,700 pounds; large producers delivering 11,700 pounds or more.

2/ Low testing producers delivered milk averaging 3.5 percent butterfat or less; high testing producers delivered milk averaging more than 3.5 percent butterfat.

3/ Variability measured by the coefficient of variation: Even producers having a coefficient of variation of 7.70 percent and under; uneven producers more than 7.70 percent.

Composite Samples

Composite samples of 10 days' deliveries were taken for 96 producers and the average of the tests of the 3 composites for the month were compared with the weighted average of daily tests. The average difference was -0.039. The standard deviation of the individual differences was

0.045, from which it can be calculated that approximately 80 percent of the composite tests would be lower than the daily tests. Actually, for 78 of the 96 producers, or 81 percent, the composite average was the lower. There were 10 differences of 0.1 or more in average fat percentages. Further details of the distribution of the differences are shown in table 5.

Table 5.--Frequency distribution of the differences from the true average (the weighted average of daily tests) of the simple monthly averages: composites and daily tests

Difference 1/	Simple average of tests of composite samples		Simple average of 30 daily tests
	10-day composites	15-day composites	
	Number	Number	Number
-0.20	1	-	-
-0.16	0	1	-
-0.15	1	3	-
-0.14	0	2	-
-0.13	0	4	-
-0.12	3	5	-
-0.11	0	2	-
-0.10	5	6	-
-0.09	1	8	-
-0.08	4	12	-
-0.07	11	4	-
-0.06	12	10	-
-0.05	4	7	-
-0.04	11	6	-
-0.03	4	5	-
-0.02	11	9	3
-0.01	10	1	25
0.00	7	4	129
+0.01	2	1	24
+0.02	1	3	9
+0.03	2	1	3
+0.04	3	0	0
+0.05	1	2	1
+0.06	1	0	-
+0.08	1	1	-
+0.17	-	1	-
Total	96	98	194

1/ Specified average minus weighted average of daily tests.

For the 98 other producers, composite samples of 15 days' deliveries were taken. The average difference in fat percentages between the averages of composite tests and those of daily tests in this series was -0.060, and the standard deviation of the individual differences was 0.053. There

were 86 averages of composite tests (or 88 percent) lower than the daily tests, compared with an expected proportion of 87 percent. There were 24 differences of 0.1 or more (table 5). These results are in accord with most earlier published work so far as comparisons can be made (table 1). For the average producer the total pounds of butterfat for the month, compared with the total pounds computed from the true monthly average test, were 5.0 pounds less when computed from the average of the 15-day composite tests and 3.3 pounds less when computed from the average of the 10-day composite tests.

Comparability of Fresh and Composite Samples

The two most important statistical properties of results of any sampling routine are: (1) The average of sample means, and (2) the dispersion or variability of sample means. It was found above that the averages of fresh tests fluctuated within narrow limits around the true monthly average tests, and that the composite samples were biased, averaging significantly less than the true tests. The dispersion or variability of composite samples was low--the variance of 10-day composites was about equal to that of the averages of 12 fresh samples (see the lower part of figure 2), and the 15-day composite was about equal to 10 fresh samples.

Because of the bias in tests of composite samples, a somewhat greater variability could be tolerated in fresh than in composite samples. There is no objective measure of how much variability is equivalent to a given amount of bias. Certain conditions can be set up, and the number of fresh samples that will equal the composite samples under these conditions can be determined. For example, the number of fresh tests that will give a weighted average of low tests equal to the weighted average of the low composite tests, can be calculated. This condition is satisfied by 5 fresh samples when compared with 10-day composites, and by 3 fresh samples when compared with 15-day composites.

This comparison between 5 fresh samples and 10-day composites is shown in the upper section of figure 2. With this number of fresh tests, there may be some objection to the excess number of large negative differences (greater than about -0.12) for fresh samples. For about 6 producers in 100, monthly average butterfat percentages based on 5 fresh samples could be expected to be low by more than -0.12, while only about 3 producers in 100 would be expected to have 10-day composites lower than -0.12.

A less liberal trade of variability for bias would result by taking the number of fresh samples for which one standard error will include an equal proportion of composite tests. On this basis, 8 fresh samples are the equivalent of 10-day composites, and 5 fresh samples are the equivalent of 15-day composites.

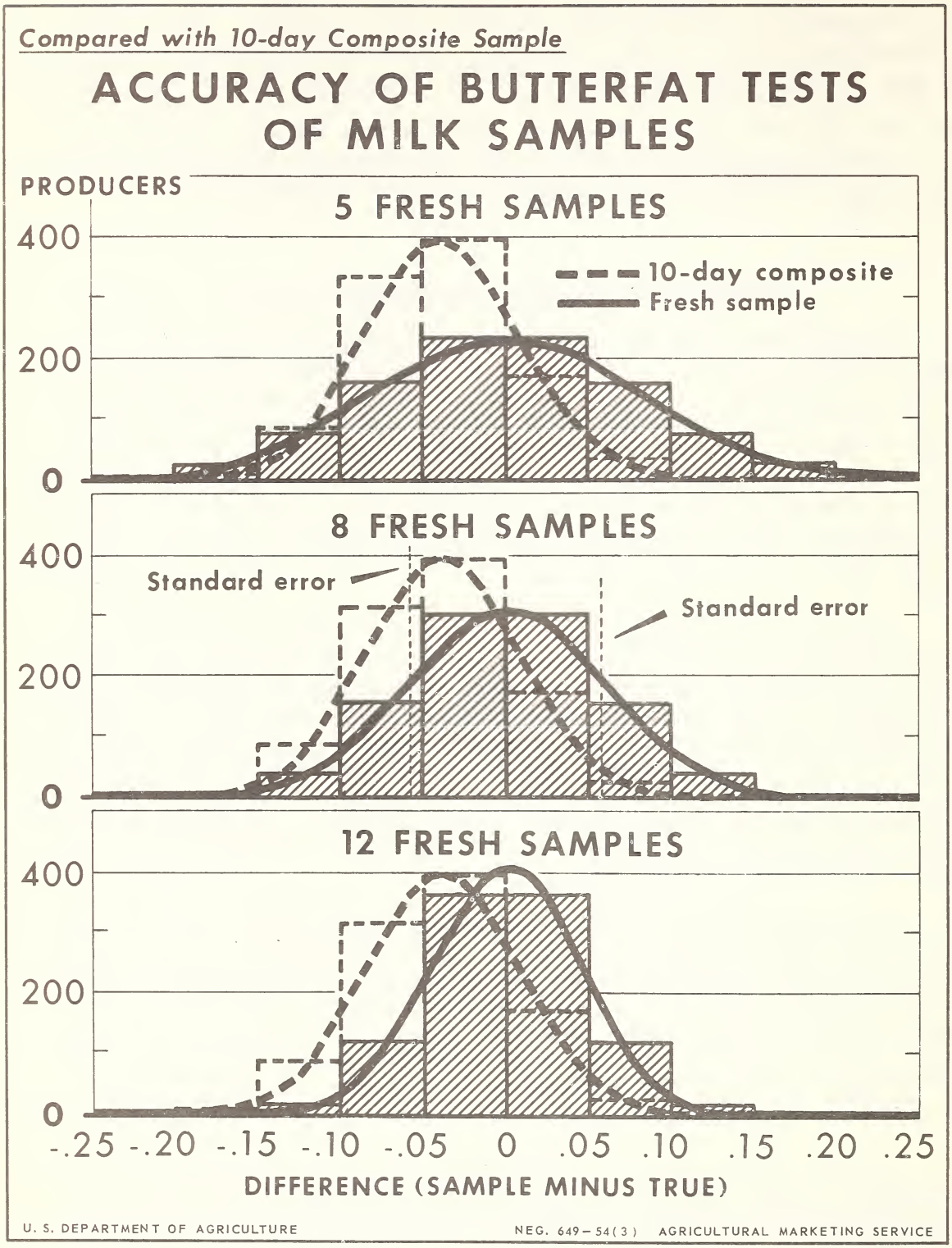

Figure 2.--Comparison of 10-day composite samples and 5, 8, and 12 fresh samples. The bars show the expected distributions of differences from true averages of 1,000 producers by class interval of 0.0-0.05, 0.05-0.10, etc. The curves show the proportionate distributions of differences of any given size for an unlimited number of producers.

This comparison of fresh and composite samples is illustrated in the center section of figure 2. The limits of one standard error for the average test of 5 fresh samples are shown. It is apparent that the area under the composite curve within these limits is about equal to the area under the curve for fresh samples. The composite samples still give more low than high tests, but the number of producers whose test would be below -0.18 would be about 12 in 10,000 for the test based on 8 fresh samples and about 9 in 10,000 for the test based on 10-day composite samples. This difference is so small that 8 fresh samples probably could be considered to be equal or superior to the 10-day composite, from the standpoint of dispersion or variability.

Simple vs. Weighted Averages of Daily Tests

In the Introduction the true average percentage of butterfat in a month's delivery of milk was defined as the average of daily percentages weighted by the quantity of milk delivered daily. A weighted average might differ appreciably from a simple average if there were (1) a close relationship between day-to-day changes in quantities of milk delivered and percentages of butterfat, and (2) a substantial variation in daily deliveries. In order to determine the importance of this consideration, the simple average of 30 daily butterfat percentages was compared with the weighted average for each of the 194 producers. The average difference was 0.0013, the simple average being the higher. For 129 in 194 producers, or about 7 in 10, the difference was less than 0.005, or one-twentieth of a "point" of butterfat.

CONCLUSIONS

Results of the study show how closely the monthly weighted average butterfat percentages of producers in this study were represented by averages of 10- and 15-day composite samples, and by samples taken and tested on any number of days chosen at random during the month. About 2 in 3 single daily tests were within 0.2 of individual producers' monthly weighted average butterfat percentage. The range of variation was only half as great when 4 daily tests were averaged. Taking one additional test during the month increased the accuracy of the results more than did allowances for temperature on the day preceding each test, for day of the week, or for week of the month. Tests of some producers were more accurate than tests of others: 3 samples gave results for producers who delivered more than 11,700 pounds of milk a month as accurate as 5 samples for producers who delivered less than that amount; 4 samples of milk testing under 3.5 percent were equivalent to 5 samples of higher testing milk; and 3 samples of deliveries that did not vary greatly from day to day were equivalent to 5 samples of deliveries that did vary greatly.

Composite samples gave results which averaged significantly lower than the monthly weighted average butterfat percentages, the differences

averaging 0.039 for 10-day composites and 0.060 for 15-day composites. The number of daily samples that can be considered as satisfactory as 10-day or 15-day composites depends upon the criteria of acceptability to plants and their patrons, but it appears to lie between 5 and 12 daily samples in comparison with 10-day composites, and between 3 and 10 daily samples in comparison with 15-day composites.

LITERATURE CITED

(1) Campbell, H. C., Jaggard, G., and Morris, D.
1931. Accuracy of the Composite Test. The Milk Dealer. 21 (3): 42.

(2) Cooke, W. W., and Hills, J. L.
1891. Creamery Work. Vt. Agr. Expt. Sta. An. Rpt., pp. 59-69.

(3) England, C. W., and D'Ambrogi, G. D.
1937. Some Factors Affecting the Accuracy of the Babcock Test on Composite Samples of Milk. Md. Agr. Expt. Sta. Bull. 413.

(4) Erdmann, H. H.
1950. A Statistical Analysis of Butterfat Tests of Milk Based on Random Fresh Milk Sampling. Federal Milk Market Administrator. Chicago, Ill.

(5) Farrington, E. H.
1891. Composite Milk Samples Tested for Butterfat. Ill. Agr. Expt. Sta. Bull. 16.

(6) Hunziker, O. F.
1914. Report of the Dairy Husbandry Department. Ind. Agr. Expt. Sta. An. Rpt. 27, pp. 37-47.

(7) Judkins, H. F.
1921. The Sampling of Milk. Jour. Dairy Sci. 4 (4): 350.

(8) Kent, F. L.
1902. Testing Milk and Cream. Oreg. Agr. Expt. Sta. Bull. 70.

(9) Lucas, P. S.
1938. Factors Involved in Accuracy of Testing Milk Samples: Accuracy of Composite Samples. Mich. Agr. Expt. Sta. Tech. Bull. 158.

(10) Marquardt, J. C.
1950. 1950 Presents the Weight Tank Problem. Amer. Butter and Cheese Rev. Vol. 12, p. 18.

(11) Meade, D., and Leckie, J. N.
1936. To What Extent do Tests From Composite Samples and Fresh Samples of Milk Agree? Milk Plant Monthly. 25 (8): 28-30.

(12) Monroe, C. F.
1929. Accuracy of Composite Milk Samples. Ohio Agr. Expt. Sta. Bull. 446.

(13) Sanmann, F. P., and Overman, O. R.
1926. Periodic Sampling Compared With Composite Sampling and the True Average Tests of Monthly Milk Deliveries. Creamery and Milk Plant Monthly. 15 (7): 39-41.

(14) Schwarzkopf, V.
1950. Weighing, Sampling, and Testing. Amer. Milk Rev., Vol. 12, No. 8.

(15) Teal, E.
1948. Five Fresh Samples as a Basis of Fat Payment to Producers. Proceedings Third Annual Meeting. Midwest. Milk Marketing Conf. Proc. 3 (1948): 35.

(16) Tracy, P. H., and Tuckey, S. L.
1939. Accuracy of Methods of Sampling Milk Deliveries at Milk Plants. Ill. Agr. Expt. Sta. Bull. 459.

(17) Wilster, G. H., and Robichaux, R. P.
1940. Sampling, Preserving, and Testing Milk. Oreg. Agr. Expt. Sta. Bull. 383.

(18) Winfrey, J. W.
1949. The Results of Adding Mechanical Agitators to Weigh Tanks. Midwest. Milk Marketing Conf., Madison, Wis. Proc., April 22-23.

APPENDIX

The data collected for this study are based on samples taken daily from milk delivered by 194 producers at a large milk plant in St. Louis. All sampling and testing was done under the supervision of J. E. Edmonson, Department of Dairy Husbandry, University of Missouri. Testing methods were those set forth in Standard Methods for the Examination of Dairy Products. 4/ All glassware used in testing was officially calibrated.

Samples were taken from a single compartment dump tank without stirring. This tank had been checked as to the accuracy of samples before and after stirring the milk, and was found to do a satisfactory job of agitation during the normal process of dumping. The tank was square, with rounded corners, about 2 feet deep, and with baffles through which the milk fell.

Samples of about one-half pint were taken from each producer's milk. The samples were taken in half pint standard glass milk bottles, and immediately capped with a plain bottle cap on which the producer's number was written. The samples were then taken to the Market Administrator's laboratory. Composite samples were made from the fresh milk after thorough agitation, and samples were pipetted for the fresh milk tests. The composite samples were kept in rubber-stoppered composite sample bottles. They were preserved with bichloride of mercury tablets and held in a refrigerator at about 48 to 50 degrees F. Temperatures of all samples were brought to about 70 degrees F. by being allowed to stand at room temperature for a half hour to an hour before they were pipetted.

As stated above, the Babcock test procedure specified by Standard Methods was used. The centrifuge was electrically driven and heated. Its speed was checked and found to be correct. The test bottles were graduated by tenths of 1 percent to 8 percent, and readings were made to the nearest tenth of 1 percent.

Three testers were employed during the experimental period. Two of the men did most of the work, and a third gave occasional assistance. The degree of uniformity among the testers was excellent. Each had had 3 to 15 years' experience in testing for butterfat. Mr. Edmonson made spot checks of their work at different times during the month. His tests were run separately after the regular testers had completed their day's work, and the results compared very closely to the original tests.

The environmental temperature measurement which was related to each butterfat test was the mean temperature of the day preceding the day on which the sample was taken, as reported by the United States Weather Bureau at St. Louis.

4/ Published by American Public Health Association, New York, and Association of Official Agricultural Chemists, Washington, D. C., 1948.

For each of the 194 producers a simple average of the 30 daily tests was computed and also an average weighted by the quantity of milk delivered daily during the 30 days. The differences between the simple and weighted averages for each producer were tabulated (table 5). Simple averages of the three 10-day composite tests of each of 96 producers and for the two 15-day composites of each of 98 producers were also computed. The differences between the averages of composite tests and the weighted averages of daily tests were also tabulated (table 5). Averages and standard deviations were computed for each distribution of differences.

For convenience in dividing the producers into 6 and 8 groups of equal size and the time period into 4 weeks, the tests on April 1 and April 30 and those for 2 producers selected at random were omitted from most of the analyses. For each of 192 producers a tabulation was made of the differences between the daily tests and the weighted average of tests for 28 days (fig. 1).

The sources of variation in butterfat tests and their significance were studied through the use of analysis of variance. The total variation was divided into the portion that was due to the differences between each producer's average and the grand average for 192 producers (among producers) and the portion that was due to the differences between each producer's daily tests and his individual monthly average (within producers). The variation within producers was in turn further divided into parts attributable to: (1) differences among average butterfat tests for days grouped according to temperature; (2) differences among the weekly averages; (3) differences among averages for each day of the week; and (4) differences among averages for each of 28 days (table 6).

The average variation of each producer's daily tests about his 28-day average (within-producer variance) was used as the basis of estimating expected differences from the true mean for averages computed from specified numbers of daily tests (table 2).

Average within-producer variances were also computed and used to estimate expected differences for producers grouped according to size of delivery, average butterfat test, and variation in milk deliveries (tables 3-4 and 8-10).

The several estimates of variances connected with known sources were compared with the unexplained random variations in order to determine the significance of the explained variations (F-tests). For some of the factors such as week and day of the week the part of the variability in butterfat tests that was correlated with variability in environmental temperatures was eliminated by the method of analysis of covariance. The variances remaining after corrections for covariance were further tested for significance (table 7).

Table 6.—Analysis of variance of daily tests of butterfat in milk delivered by producers

Source of variation	Degrees of freedom	Sum of squares	Variance	Standard deviation
Among producers	191	546.98	2.8638	-
Within producers (Total)	5,184	198.43	.0383	0.1957
Among temperature groups	3	3.38	1.1267	-
Within producers less temperature	5,181	195.05	.0376	0.1939
Among weeks	3	1.12	.3733	-
Within producers less weeks	5,181	197.31	.0381	0.1952
Among days of week	6	1.34	.2233	-
Within producers less days of week	5,178	197.09	.0381	0.1952
Among days	27	6.68	.2474	-
Residual (Error)	5,157	191.75	.0372	0.1929
Total	5,375	745.41	-	-

Table 7.—Variation of average tests among weeks, corrected for the regression of butterfat tests on environmental temperature

Source of variation	Degrees of freedom	Reduced sum of squares	Reduced variance	F 1/
Among weeks	3	0.17	0.0567	2/ 1.79
Error	3,437	108.47	0.0316	-
Error and weeks	3,440	108.64	-	-

1/ F is the ratio of the two estimates of reduced variance.
2/ Not significant.

Table 8.--Analysis of variance of daily butterfat tests of producers grouped according to quantities of milk delivered

Milk delivered	Source of variation	Degrees of freedom	Sum of squares	Variance	Standard deviation
Pounds					
1,200-4,799	Among producers	23	159.97	6.9552	-
	Within producers	648	30.47	.0470	0.2168
	Total	671	190.44	-	-
4,800-6,179	Among producers	23	83.91	3.6483	-
	Within producers	648	29.68	.0458	0.2140
	Total	671	113.59	-	-
6,180-7,249	Among producers	23	35.02	1.5226	-
	Within producers	648	27.21	.0420	0.2049
	Total	671	62.23	-	-
7,250-8,299	Among producers	23	71.55	3.1109	-
	Within producers	648	25.97	.0401	0.2002
	Total	671	97.52	-	-
8,300-9,999	Among producers	23	45.99	1.9996	-
	Within producers	648	21.87	.0338	0.1838
	Total	671	67.86	-	-
10,000-11,699	Among producers	23	31.52	1.3704	-
	Within producers	648	28.19	.0435	0.2086
	Total	671	59.71	-	-
11,700-14,199	Among producers	23	38.56	1.6765	-
	Within producers	648	18.15	.0280	0.1673
	Total	671	56.71	-	-
14,200-19,882	Among producers	23	23.38	1.0165	-
	Within producers	648	16.89	.0261	0.1616
	Total	671	40.27	-	-

Table 9.--Analysis of variance of daily butterfat tests of producers grouped according to monthly average butterfat test

Butterfat test	Source of variation	Degrees of freedom	Sum of squares	Variance	Standard deviation
Percent					
2.60-3.16	Among producers	23	6.84	0.2974	-
	Within producers	648	25.94	0.0400	0.2000
	Total	671	32.78	-	-
3.17-3.27	Among producers	23	.59	0.0257	-
	Within producers	648	18.16	0.0280	0.1673
	Total	671	18.75	-	-
3.28-3.34	Among producers	23	.32	0.0139	-
	Within producers	648	18.72	0.0288	0.1697
	Total	671	19.04	-	-
3.35-3.41	Among producers	23	.35	0.0152	-
	Within producers	648	20.53	0.0317	0.1780
	Total	671	20.88	-	-
3.42-3.51	Among producers	23	.94	0.0409	-
	Within producers	648	22.10	0.0341	0.1847
	Total	671	23.04	-	-
3.52-3.61	Among producers	23	.87	0.0378	-
	Within producers	648	28.83	0.0445	0.2110
	Total	671	29.70	-	-
3.62-3.76	Among producers	23	1.24	0.0539	-
	Within producers	648	24.80	0.0383	0.1957
	Total	671	26.04	-	-
3.77-4.99	Among producers	23	81.11	3.5265	-
	Within producers	648	39.35	0.0607	0.2464
	Total	671	120.46	-	-

Table 10.—Analysis of variance of daily butterfat tests of producers grouped according to variability of milk deliveries

Coefficient of variation of milk deliveries	Source of variation	Degrees of freedom	Sum of squares	Variance	Standard deviation
Percent					
2.00-4.59	Among producers	31	70.34	2.2690	-
	Within producers	864	21.41	.0248	0.1575
	Total	895	91.75	-	-
4.60-6.12	Among producers	31	47.89	1.5448	-
	Within producers	864	22.90	.0265	0.1628
	Total	895	70.79	-	-
6.13-7.69	Among producers	31	45.05	1.4532	-
	Within producers	864	28.90	.0334	0.1828
	Total	895	73.95	-	-
7.70-10.49	Among producers	31	83.89	2.7061	-
	Within producers	864	41.00	.0475	0.2179
	Total	895	124.89	-	-
10.50-14.19	Among producers	31	150.15	4.8435	-
	Within producers	864	41.29	.0478	0.2186
	Total	895	191.44	-	-
14.20-38.02	Among producers	31	119.66	3.8600	-
	Within producers	864	42.93	.0497	0.2229
	Total	895	162.59	-	-

U. S. GOVERNMENT PRINTING OFFICE: O—1954